A Note on Bootstrap Variance Estimation

by

Jun Shao
Purdue University

Technical Report #88-29

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

1

A Note on Bootstrap Variance Estimation

by

Jun Shao
Purdue University

Technical Report #88-29

Department of Statistics
Purdue University

June, 1988

A NOTE ON BOOTSTRAP VARIANCE ESTIMATION

Jun Shao *
Department of Statistics, Purdue University
West Lafayette, IN 47907, U.S.A.

ABSTRACT

The bootstrap estimator of the asymptotic covariance matrix of a function of sample means or sample quantiles is inconsistent in some situations. A modified bootstrap estimator is proposed and shown to be consistent under weak conditions. A simulation study shows that in terms of finite-sample performance, the improvement of this modification is substantial. The computation of ~~our~~ this modified bootstrap estimator is much easier and cheaper than that of the estimator based on the quantiles of the bootstrap distribution. ~~We show~~ It is shown by simulation that with the same number of bootstrap replicates (in bootstrap Monte Carlo approximation), the modified bootstrap estimator is more accurate than the estimator based on the interquartile range of the bootstrap distribution. Keyword:

Key words. Asymptotic variance, consistency, sample mean, sample quantile, truncation. (KF)

* The research of this author was partially supported by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grant DMS-8606964, DMS-8702620 at Purdue University.

1. INTRODUCTION

Let μ be an unknown characteristic of a population distribution F. We focus on the following two cases which are frequently encountered in practice: (i) F is k-variate and $\mu = \int x dF$, the mean of F; (ii) F is univariate and $\mu = (\, Q(p_1),..., Q(p_k)\,)'$, where $Q(p_j)$ is the p_j-quantile of F. The quantity of interest is $\theta = g(\mu)$, where g is a fixed function from $\mathbf{R}^k$ to $\mathbf{R}^m$.

Let $X_1,..., X_n$ be independent and identically distributed (i.i.d.) samples from F. A point estimator of θ in case (i) is $\hat{\theta} = g(\bar{X})$, where $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ is the sample mean. For case (ii), let $\hat{Q}(p_j)$ be the sample p_j-quantile based on $X_1,..., X_n$ and $\hat{Q} = (\, \hat{Q}(p_1),..., \hat{Q}(p_k)\,)'$. A point estimator of θ is then $\hat{\theta} = g(\hat{Q})$.

It is well known that under reasonable conditions $n^{1/2}(\hat{\theta}-\theta)$ converges in law (as the sample size $n \to \infty$) to an m-variate normal distribution with mean zero and covariance matrix Σ. The Σ is called the asymptotic covariance matrix of $\hat{\theta}$ and is usually unknown. For assessing the accuracy of the point estimator $\hat{\theta}$, we need an estimator of Σ. Obtaining a good estimator of Σ is also crucial for making other statistical inferences such as testing hypothesis and setting confidence region for θ.

Efron (1979) introduced a bootstrap method for variance estimation. Let $X_1^*,..., X_n^*$ be i.i.d. samples from $\{ X_1,..., X_n \}$, $\bar{X}^* = n^{-1}\sum_{i=1}^n X_i^*$ and $\hat{Q}^*$ be the k-vector of sample quantiles based on $X_1^*,..., X_n^*$. Let $\hat{\theta}^* = g(\bar{X}^*)$ if $\hat{\theta} = g(\bar{X})$ and $\hat{\theta}^* = g(\hat{Q}^*)$ if $\hat{\theta} = g(\hat{Q})$. The bootstrap estimator of the asymptotic covariance matrix Σ of $\hat{\theta}$ is then

$$\hat{\Sigma}_b = nVar_*(\hat{\theta}^*) = nE_*(\hat{\theta}^* - E_*\hat{\theta}^*)(\hat{\theta}^* - E_*\hat{\theta}^*)', \tag{1.1}$$

where E_* and Var_* are the expectation and variance taken under the bootstrap distribution.

An essential theoretical justification of a variance estimator is its consistency. When g is the identity function, the bootstrap estimator $\hat{\Sigma}_b$ is consistent. For the case of $\hat{\theta} = \bar{X}$,

$$\hat{\Sigma}_b = n^{-1}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' \to \Sigma \quad a.s.$$

according to the strong law of large numbers. For the case of $\hat{\theta} = \hat{Q}$, the consistency of $\hat{\Sigma}_b$

was proved by Babu (1986) under some conditions (see Theorem 2).

However, even for smooth differentiable function g, the consistency of $\hat{\Sigma}_b$ is not guaranteed. A counter-example is given in Section 2. To circumvent the inconsistency of the bootstrap variance estimator, we propose a modified bootstrap variance estimator. Description of this modification is given in Section 2. The consistency of the modified bootstrap variance estimator for the cases of functions of sample means and sample quantiles is established (Section 2.3). Variance estimators based on the quantiles of the bootstrap distribution, such as a multiple of the interquartile range of the bootstrap distribution, are also consistent. But the computation of our modified bootstrap estimator is much easier and cheaper than that of the bootstrap quantiles. In Section 3, simulation results show that in the case of estimating variances of functions of sample median, the modified bootstrap estimator significantly outperforms the original bootstrap estimator and the estimator based on interquartile range of the bootstrap distribution in terms of finite-sample sampling properties.

2. THE MODIFIED BOOTSTRAP ESTIMATOR

2.1. *A Counter-example*

The following example shows that the bootstrap estimator (1.1) may be inconsistent.

We consider the univariate case. Let F be a univariate distribution function satisfying $F(x) = 1-x^{-h}$ if $x>10$ and $F(x) = |x|^{-h}$ if $x<-10$, where h is a constant. Thus, F has finite sth moment for any $s<h$. In particular, F has finite second moment if $h>2$. Let $t>h$ be a constant and $g(x) = \exp(x^t)$. Following the proof in Ghosh et al. (1984, Example), the bootstrap variance estimator for the case where $\hat{\theta}$ is either $g(\bar{X})$ or $g(\hat{Q})$ (with $0<p<1$) is inconsistent if

$$n^{-n+1}[g(X_{(n)})]^2 \to \infty \quad a.s., \tag{2.1}$$

where $X_{(n)} = \max(X_1, \ldots, X_n)$. In fact, under (2.1), $nVar_*(\hat{\theta}^*) \to \infty$ $a.s.$

To show (2.1), note that for any $M>0$,

$$P\{\, n^{-n+1}[g(X_{(n)})]^2 < M \,\} \le P\{\, X_{(n)} < [\log(M^{1/2} n^{(n-1)/2})]^{1/t} \,\}$$

$$= \{1-[\log(M^{1/2} n^{(n-1)/2})]^{-h/t}\}^n \le \exp\{-n[\log(M^{1/2} n^{(n-1)/2})]^{-h/t}\} \le n^{-2}$$

for large n. Thus, (2.1) follows from the Borel-Cantelli lemma.

2.2. *A Modification*

The above example shows that the bootstrap variance estimator may diverge to infinity while the asymptotic variance of $\hat{\theta}$ is finite. The inconsistency of the bootstrap estimator is caused by the fact that $\|\hat{\theta}^* - \hat{\theta}\|$ may take some exceptionally large values, where $\|x\| = (x'x)^{1/2}$ for any vector x. A remedy is to truncate $\hat{\theta}^* - \hat{\theta}$ at some value. Throughout the paper, the jth components of $\hat{\theta}^*$ and $\hat{\theta}$ are denoted by $\hat{\theta}_j^*$ and $\hat{\theta}_j$, respectively. Let $\tau(X) = \tau(X_1, ..., X_n)$ be a k-vector of functions of data satisfying

$$\tau_j \ge c_0 \quad \text{and} \quad \tau_j = O(1) \ \ a.s. \quad j=1,...,k, \tag{2.2}$$

where τ_j is the jth component of $\tau(X)$ and c_0 is a fixed constant. A modified bootstrap estimator of Σ is

$$\hat{\Sigma}_a = nVar_*(\Delta^*), \tag{2.3}$$

where $\Delta^* = (\Delta_1^*, ..., \Delta_k^*)'$ and

$$\Delta_j^* = \begin{cases} \tau_j & \text{if } \hat{\theta}_j^* - \hat{\theta}_j > \tau_j \\ \hat{\theta}_j^* - \hat{\theta}_j & \text{if } |\hat{\theta}_j^* - \hat{\theta}_j| \le \tau_j \\ -\tau_j & \text{if } \hat{\theta}_j^* - \hat{\theta}_j < -\tau_j \end{cases} \tag{2.4}$$

In the following we establish the consistency of the modified bootstrap estimator $\hat{\Sigma}_a$ under some weak conditions. Choices of the function $\tau(X)$ are discussed in Section 2.4.

2.3. *Consistency of the Modified Bootstrap Estimator*

Let F be a k-variate distribution function, $\mu = EX_1$, $\theta = g(\mu)$, $\hat{\theta} = g(\bar{X})$ and ∇g be the gradient of g. If $E\|X_1\|^2 < \infty$ and ∇g is continuous in a neighborhood of μ, then as $n \to \infty$,

$$n^{1/2}(\hat{\theta}-\theta) \to Z \quad \text{in law}, \tag{2.5}$$

where Z has an m-variate normal distribution with mean zero and covariance matrix

$$\Sigma = \nabla g(\mu) Var(X_1)(\nabla g(\mu))'.$$

The proof of the following theorem is given in the Appendix.

Theorem 1. Assume that $E\|X_1\|^2 < \infty$ and g is continuously differentiable in a neighborhood of μ. Then the modified bootstrap estimator $\hat{\Sigma}_a$ (defined in (2.2)-(2.4)) is consistent, i.e., as $n \to \infty$,

$$\hat{\Sigma}_a \to \Sigma \quad a.s.$$

For the sample quantiles, we consider univariate F. Let the jth component of μ be $Q(p_j)$ (p_j-quantile of F), $0<p_j<1$, $j=1,\ldots,k$, $\theta = g(\mu)$, $\hat{\theta} = g(\hat{Q})$, and $\hat{\Sigma}_a$ be defined in (2.2)-(2.4). It is well known that $n^{1/2}(\hat{\theta}-\theta)$ converges in law to an m-variate normal distribution with mean zero and covariance matrix

$$\Sigma = \nabla g(\mu)\Lambda(\nabla g(\mu))', \tag{2.6}$$

where Λ is a $k \times k$ symmetric matrix whose (i,j)th element is

$$\lambda_{ij} = p_i(1-p_j)/[f(Q(p_i))f(Q(p_j))], \quad 1 \le i \le j \le k,$$

$f(Q(p_i))$ is the derivative of F at $Q(p_i)$ and is assumed to be positive.

We have the following result (the proof is in the Appendix).

Theorem 2. Assume that F is differentiable at $Q(p_j)$ with $f(Q(p_j))>0$ and $0<p_j<1$, $j=1,\ldots,k$, where f is the derivative of F. Assume also that $E[\log(1+|X_1|)] < \infty$ and g is continuously differentiable in a neighborhood of $\mu = (Q(p_1),\ldots,Q(p_k))'$. Then

$$\hat{\Sigma}_a \to \Sigma \quad a.s.$$

2.4. *Some Practical Issues*

The modified bootstrap estimator $\hat{\Sigma}_a$ is consistent (under the weak conditions in Theorems 1 and 2) for any function $\tau(X)$ satisfying (2.2). Two choices of the function $\tau(X)$ for practical uses are suggested as follows.

(1) $\tau_j \equiv$ a constant. This can be used when one has some rough information about the asymptotic variance of $\hat{\theta}_j$. For example, the asymptotic variance is unknown but bounded by a positive constant C. Then τ_j can be chosen to be any constant $\tau > C^{1/2}$.

(2) $\tau_j = \max(\rho|\hat{\theta}_j|, c_0)$ for a small positive constant c_0 and a positive constant ρ. Clearly this τ_j satisfies (2.2) if $\hat{\theta}$ is strongly consistent. The small constant c_0 is used to prevent τ_j approaching zero. With this choice of τ_j, $|\hat{\theta}_j^* - \hat{\theta}_j|$ is replaced by τ_j when the ratio $\hat{\theta}_j^*/\hat{\theta}_j$ differs from one by more than $\pm 100\rho\%$. A simulation study of the performance of $\hat{\Sigma}_a$ with this choice of τ_j is given in Section 3.

For numerical evaluation of the bootstrap estimator, Efron (1979) proposed the use of the Monte Carlo approximation. The same idea can be used here for the evaluation of the modified bootstrap estimator. That is, we generate i.i.d. samples $X_1^{*b}, \ldots, X_n^{*b}$ from $\{X_1, \ldots, X_n\}$, $b=1,\ldots,B$, and calculate Δ^{*b} (based on $X_1^{*b}, \ldots, X_n^{*b}$) according to (2.4). Then use

$$B^{-1}\sum_{b=1}^{B}(\Delta^{*b} - B^{-1}\sum_{b=1}^{B}\Delta^{*b})^2$$

to approximate $Var_*(\Delta^*)$.

2.5. *Comparison with the estimator based on bootstrap quantiles*

Consider the situation where θ is a scalar ($m=1$). Let α be a constant between 0 and $1/2$. Then the following estimator of the asymptotic variance of $n^{1/2}(\hat{\theta}-\theta)$ is consistent:

$$\hat{\Sigma}_q = [H^{-1}(1-\alpha) - H^{-1}(\alpha)]/[\Phi^{-1}(1-\alpha) - \Phi^{-1}(\alpha)],$$

where Φ is the standard normal distribution, $H(x) = P_*\{ n^{1/2}(\hat{\theta}^* - \hat{\theta}) \le x \}$, and $\Phi^{-1}(a)$ and $H^{-1}(a)$ are the a-quantile of Φ and H, respectively. An example is $\alpha=1/4$ and $\hat{\Sigma}_q$ is a multiple of the interquartile range of the bootstrap distribution H.

Although $\hat{\Sigma}_q$ is consistent and therefore asymptotically equivalent to the modified bootstrap estimator $\hat{\Sigma}_a$, the computation of $\hat{\Sigma}_a$ for any fixed sample size is easier and cheaper than that of $\hat{\Sigma}_q$, since the former involves the computation of the second order moment of the bootstrap distribution H whereas the latter involves the computation of the quantiles of H. Usually $\hat{\Sigma}_a$ and $\hat{\Sigma}_q$ have to be approximated by Monte Carlo (see Section 2.4). Obtaining an accurate Monte Carlo approximation of the second order moment of the bootstrap distribution H is much easier than obtaining an accurate Monte Carlo approximation of the quantiles of H. It was shown (Efron, 1987, Section 9) that the Monte Carlo approximation of the second order moment of H usually requires 100~200 bootstrap replications. On the other hand, the Monte Carlo approximation of a quantile of H is more costly, requiring 1000~2000 bootstrap replications. The amount of computation required for $\hat{\Sigma}_q$ is at least 10 times as much as that for $\hat{\Sigma}_a$.

For the same bootstrap replication size B, $\hat{\Sigma}_q$ is much less accurate than $\hat{\Sigma}_a$ and is also less accurate than $\hat{\Sigma}_b$ when $\hat{\Sigma}_b$ is consistent. This is shown in the following simulation study.

3. A SIMULATION STUDY

In this section we study by simulation the finite-sample sampling properties of the modified bootstrap estimator, the original bootstrap estimator and the estimator based on bootstrap interquartile range in the case of estimating the asymptotic variances of functions of sample median.

Let $\hat{Q}$ be the sample median based on n=36 i.i.d. samples from a distribution F and $\hat{\theta}=g(\hat{Q})$. Three functions g are considered: (i) $g(x)=x$; (ii) $g(x)=x^2/4$; (iii) $g(x)=e^x/4$. Two distributions F under consideration are: (i) normal distribution with median (mean) 1.5 and standard deviation 2; (ii) Cauchy distribution with median 1.5 and scale parameter 2.

The function $\tau(X)$ for the modified bootstrap estimator is chosen to be $\max(\tfrac{1}{2}|\hat{\theta}|, 0.05)$. For the evaluation of the three bootstrap estimators, Monte Carlo approximation of size B=500 is used (see Section 2.4).

Table 1 reports the root mean squared errors (rmse) and the biases of the three bootstrap estimators. The asymptotic variances (denoted by σ^2) are included. All the results are based on 2000 simulations on a VAX 11/780 at Purdue University. The IMSL subroutines are used for generating random numbers.

We summarize the simulation results as follows.

(1) *Overall.* All three bootstrap variance estimators are up-ward biased. The modified bootstrap estimator reduces the bias considerably. In terms of the rmse, the modified bootstrap significantly out-performs the original bootstrap and the bootstrap interquartile range. The ratio of the rmse of the modified bootstrap estimator to the rmse of the original bootstrap estimator (or the bootstrap interquartile range), denoted by R, is shown in Table 1.

(2) *The modified bootstrap and the original bootstrap.* The improvement of the modified bootstrap over the original bootstrap is larger if the distribution F has heavier tails and/or the function $g(x)$ has a faster rate of divergence (as $|x| \to \infty$). This indicates that even if the original bootstrap estimator is consistent, the modified bootstrap estimator may have a faster convergence rate.

(3) *The modified bootstrap and the interquartile range.* With the same bootstrap replication number B=500, the modified bootstrap is much more efficient than the bootstrap interquartile range: the ratio R is usually about 0.5–0.6. In fact, the bootstrap interquartile range is also not as good as the original bootstrap estimator in the case where the original bootstrap estimator is consistent.

(4) *The effects of distribution tails and function g.* The case of F = Cauchy distribution and $g(x) = e^x/4$ is an exceptional case: the original bootstrap estimator is inconsistent (diverges to infinity) and the biases and rmse of the other two estimators are also very large. This indicates that although the modified bootstrap and bootstrap interquartile range estimators are consistent, the sample size n=36 is not large enough when the distribution F has heavy tails and $g(x)$ diverges to infinity at a very fast rate. However, the result in Table 1 still clearly shows that the modified bootstrap estimator is much better.

APPENDIX

Proof of Theorem 1. From Bickel and Freedman (1981), for almost all $X_1, X_2, \ldots$, the conditional distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ converges to the distribution of Z (given in (2.5)). Let $X_1, X_2, \ldots$ be a fixed sequence such that (2.2) holds and the conditional distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ converges to the distribution of Z. Let P_* be the bootstrap conditional probability and λ be an arbitrary nonzero m-vector. For any fixed $t > 0$,

$$| P_*\{ n\lambda'(\hat{\theta}^* - \hat{\theta})(\hat{\theta}^* - \hat{\theta})'\lambda < t \} - P_*\{ n\lambda'(\Delta^* \Delta^{*\prime})\lambda < t \} |$$

$$\leq 1 - P_*\{ |\hat{\theta}_j^* - \hat{\theta}_j| < \tau_j, j=1,\ldots,k, \} \to 0$$

as $n \to \infty$. Therefore the conditional distribution of $n(\Delta^* \Delta^{*\prime})$ converges to the distribution of ZZ'. It remains to show that there is a constant $\delta > 0$ such that

$$E_*(n^{1/2} \| \Delta^* \|)^{2+\delta} = O(1) \quad a.s. \tag{A1}$$

We now show that (A1) holds with $\delta = 2$. Since ∇g is continuous in a neighborhood of μ, there are positive constants η and M such that

$$trace\{[\nabla g(x)]'[\nabla g(x)]\} \leq M \quad \text{if} \quad \|x - \mu\| \leq 2\eta.$$

By the strong law of large numbers, almost surely,

$$\bar{X} \to \mu \quad \text{and} \quad n^{-1}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' \to Var(X_1). \tag{A2}$$

Let X_{ij} and $\bar{X}_j$ be the jth components of X_i and $\bar{X}$, respectively. By the Marcinkiewicz's strong law of large numbers, almost surely,

$$n^{-2}\sum_{i=1}^n (X_{ij} - \bar{X}_j)^4 \leq 16 n^{-2}\sum_{i=1}^n (X_{ij} - EX_{ij})^4 \to 0 \quad \text{for all } j=1,\ldots,k. \tag{A3}$$

Let $X_1, X_2, \ldots$ be a sequence such that (2.2), (A2) and (A3) hold. Then $\|\bar{X} - \mu\| \leq \eta$ for large n. Let $I(A)$ be the indicator function of the set A. Then

$$n^2 E_* \|\Delta^*\|^4 = n^2 E_* \|\Delta^*\|^4 I(\|\bar{X}^* - \bar{X}\| \leq \eta) + n^2 E_* \|\Delta^*\|^4 I(\|\bar{X}^* - \bar{X}\| > \eta)$$

$$\leq n^2 E_* \|\hat{\theta}^* - \hat{\theta}\|^4 I(\|\bar{X}^* - \bar{X}\| \leq \eta) + \|\tau(X)\|^4 n^2 E_* I(\|\bar{X}^* - \bar{X}\| > \eta)$$

$$= n^2 E_* \|\nabla g(\xi^*)(\bar{X}^* - \bar{X})\|^4 I(\|\bar{X}^* - \bar{X}\| \leq \eta) + \|\tau(X)\|^4 n^2 E_* I(\|\bar{X}^* - \bar{X}\| > \eta) \tag{A4}$$

$$\leq M^2 n^2 E_* \|\bar{X}^* - \bar{X}\|^4 I(\|\bar{X}^* - \bar{X}\| \leq \eta) + \eta^{-4}\|\tau(X)\|^4 n^2 E_* \|\bar{X}^* - \bar{X}\|^4 \tag{A5}$$

$$\leq (M^2 + \eta^{-4}\|\tau(X)\|^4) n^2 E_* \|\bar{X}^* - \bar{X}\|^4,$$

where (A4) follows from the mean-value theorem and ξ^* is a point on the line segment between $\bar{X}^*$ and $\bar{X}$, and (A5) follows from

$$\|\xi^* - \mu\| \leq \|\bar{X} - \mu\| + \|\xi^* - \bar{X}\| \leq \eta + \|\bar{X}^* - \bar{X}\|.$$

Under (2.2), $\|\tau(X)\| = O(1)$. Hence the result follows from

$$n^2 E_* (\bar{X}_j^* - \bar{X}^*)^4 = O(1), \tag{A6}$$

where $\bar{X}_j^*$ is the jth component of $\bar{X}^*$. A straightforward calculation shows that

$$n^2 E_* (\bar{X}_j^* - \bar{X}^*)^4 = n^{-2} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^4 + 3(n^{-2} - n^{-3})[\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2]^2.$$

Hence (A6) follows from (A2)-(A3) and thus the result. □

Proof of Theorem 2. From Bickel and Freedman (1981), for almost all $X_1, X_2, \ldots$, the conditional distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ converges to the normal distribution with mean zero and covariance matrix given by (2.6). Following the same argument in the proof of Theorem 1, we only need to show (A1).

Replacing $\bar{X}^*$ and $\bar{X}$ by $\hat{Q}^*$ and $\hat{Q}$ in the proof of Theorem 1, we have

$$n^2 E_* \|\Delta^*\|^4 \leq C_1 n^2 E_* \|\hat{Q}^* - \hat{Q}\|^4, \tag{A7}$$

where C_1 is a positive constant. Then (A1) follows from (A7) and

$$n^2 E_* \|\hat{Q}^* - \hat{Q}\|^4 = O(1) \quad a.s.$$

under $E[\log(1 + |X_1|)] < \infty$ (see Babu, 1986). □

REFERENCES

Babu, G. J. (1986). A note on bootstrapping the variance of sample quantile. *Ann. Inst. Statist. Math.* **38**, Part A, 439-443.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196-1217.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**, 1-26.

Efron, B. (1987). Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **82**, 171-200.

Ghosh, M., Parr, W. C., Singh, K. and Babu, G. J. (1984). A note on bootstrapping the sample median. *Ann. Statist.* **12**, 1130-1135.

Table 1. Results of simulation comparison of the modofied bootstrap, the original bootstrap and the interquartile range estimators.

Normal distribution									
		Modified bootstrap		Original bootstrap			Interquartile range		
$g(x)$	σ^2/n	bias	rmse	bias	rmse	$R^{\ddagger}$	bias	rmse	$R^{\ddagger}$
x	0.1745	0.0108	0.0985	0.0150	0.1049	0.9390	0.0208	0.1747	0.5638
$x^2/4$	0.0982	0.0070	0.0722	0.0193	0.0921	0.7839	0.0194	0.1361	0.5305
$e^x/4$	0.2191	0.1252	0.3330	0.2211	0.5908	0.5636	0.1363	0.5724	0.5818
Cauchy distribution									
		Modified bootstrap		Original bootstrap			Interquartile range		
$g(x)$	σ^2/n	bias	rmse	bias	rmse	$R^{\ddagger}$	bias	rmse	$R^{\ddagger}$
x	0.2742	0.0617	0.1928	0.1037	0.2605	0.7401	0.0782	0.3495	0.5516
$x^2/4$	0.1542	0.0405	0.1747	0.1302	0.4320	0.4044	0.0737	0.3607	0.4843
$e^x/4$	0.3442	0.5556	1.6558	1.03×10^3	4.56×10^4	0.0000	0.8112	8.3566	0.1981

$^{\ddagger}R = \dfrac{\text{rmse of modified bootstrap}}{\text{rmse}}$

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE:

3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release, distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S): Technical Report #88-29

5. MONITORING ORGANIZATION REPORT NUMBER(S):

6a. NAME OF PERFORMING ORGANIZATION: Purdue University

6b. OFFICE SYMBOL *(If applicable)*:

7a. NAME OF MONITORING ORGANIZATION:

6c. ADDRESS *(City, State, and ZIP Code)*:
Department of Statistics
West Lafayette, IN 47907

7b. ADDRESS *(City, State, and ZIP Code)*:

8a. NAME OF FUNDING / SPONSORING ORGANIZATION: Office of Naval Research

8b. OFFICE SYMBOL *(If applicable)*:

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-88-K-0170 and NSF Grants DMS-8606964 DMS-8702620

8c. ADDRESS *(City, State, and ZIP Code)*: Arlington, VA 22217-5000

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*: A NOTE ON BOOTSTRAP VARIANCE ESTIMATION (Unclassified)

12. PERSONAL AUTHOR(S): Jun Shao

13a. TYPE OF REPORT: Technical

13b. TIME COVERED: FROM ______ TO ______

14. DATE OF REPORT *(Year, Month, Day)*: June 1988

15. PAGE COUNT: 10

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*: asymptotic variance, consistency, sample mean, sample quantile, truncation

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The bootstrap estimator of the asymptotic covariance matrix of a function of sample means or sample quantiles is inconsistent in some situations. A modified bootstrap estimator is proposed and shown to be consistent under weak conditions. A simulation study shows that in terms of finite-sample performance, the improvement of this modification is substantial. The computation of our modified bootstrap estimator is much easier and cheaper than that of the estimator based on the quantiles of the bootstrap distribution. We show by simulation that with the same number of bootstrap replicates (in bootstrap Monte Carlo approximation), the modified bootstrap estimator is more accurate than the estimator based on the interquartile range of the bootstrap distribution.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: Jun Shao

22b. TELEPHONE *(Include Area Code)*: (317) 494-6039

22c. OFFICE SYMBOL: